**ĐỀ TÀI PHƯƠNG PHÁP PHÂN LOẠI ĐIỂM TÍN DỤNG**

|  |  |  |
| --- | --- | --- |
| Id | Cite | Abstract |
|  | @article{Jakka2023ANC,  title={A novel credit scoring system in financial institutions using artificial intelligence technology},  author={Geethamanikanta Jakka and Amrutanshu Panigrahi and Abhilash Pati and Manmath Nath Das and Jyotsnarani Tripathy},  journal={Journal of Autonomous Intelligence},  year={2023}, | A novel credit scoring system in financial institutions using artificial intelligence technology In order to evaluate a person’s or a company’s creditworthiness, financial institutions must use credit scoring. Traditional credit scoring algorithms frequently rely on manual and rule-based methods, which can be tedious and inaccurate. Recent developments in artificial intelligence (AI) technology have opened up possibilities for creating more reliable and effective credit rating systems. The data are pre-processed, including scaling using the 0–1 normalization method and resolving missing values by imputation. Information gain (IG), gain ratio (GR), and chi-square are three feature selection methodologies covered in the study. While GR normalizes IG by dividing it by the total entropy of the feature, IG quantifies the reduction in total entropy by adding a new feature. Based on chi-squared statistics, the most vital traits are determined using chi-square. This research employs different ML models to develop a hybrid model for credit score prediction. The ML algorithms support vector machine (SVM), neural networks (NNs), decision trees (DTs), random forest (RF), and logistic regression (LR) classifiers are employed here for experiments along with IG, GR, and chi-square feature selection methodologies for credit prediction over Australian and German datasets. The study offers an understanding of the decision-making process for informative characteristics and the functionality of machine learning (ML) in credit prediction tasks. The empirical analysis shows that in the case of the German dataset, the DT with GR feature selection and hyperparameter optimization outperforms SVM and NN with an accuracy of 99.78%. For the Australian dataset, SVM with GR feature selection outperforms NN and DT with an accuracy of 99.98%. Keywords: credit scoring system; machine learning (ML); classification techniques; feature selection algorithms; hyperparameter optimization |
|  | @article{Chacko2023EnhancingCS,  title={Enhancing Credit Score Analysis: A Novel Approach with Random Forest and Kernel SVM},  author={Annie Chacko and John Aravindhar D and Antonidoss A},  journal={International Journal of Electronics and Communication Engineering},  year={2023}, | Enhancing Credit Score Analysis: A Novel Approach with Random Forest and Kernel SVM  Credit score analysis systematically evaluates an individual or entity’s financial history and behaviour to determine their creditworthiness. Traditional methods for credit score analysis have several challenges, such as privacy concerns, lack of flexibility, vulnerability to identity theft, limited data, and real-time analysis. To overcome these complexities, this paper proposes a novel method combining the advantages of Random Forest and kernel Support Vector Machine (SVM). The proposed method has three phases: data preprocessing, feature extraction, and classification. In the preprocessing phase, the proposed method eliminates the noise and errors from the raw data based on obtaining quality input for the analysis. In this study, Random Forest is utilized to extract the most significant features based on the domain and credit data analysis also, kernel SVM is employed for classification by analyzing the components and their impact on credit scoring. Also, the study conducted experiments on the German Credit Risk dataset. The performance evaluation of the proposed method involves analyzing its effectiveness based on evaluation metrics and comparing its performance with existing methods. The experimental results depict that the proposed method obtained better outcomes and achieved high efficiency for credit score analysis. Keywords - Credit score analysis, Random Forest, Kernel Support Vector Machine, German Credit Risk dataset, Feature extraction, Classification. |
|  | @article{Markov2022CreditSM,  title={Credit Scoring Methods: Latest Trends and Points to Consider},  author={Anton Markov and Zinaida Seleznyova and Victor A. Lapshin},  journal={The Journal of Finance and Data Science},  year={2022}, | Credit scoring methods: Latest trends and points to consider Credit risk is the most significant risk by impact for any bank and financial institution. Accurate credit risk assessment affects an organisation's balance sheet and income statement, since credit risk strategy determines pricing, and might even influence seemingly unrelated domains, e.g. marketing, and decision-making. This article aims at providing a systemic review of the most recent (2016–2021) articles, identifying trends in credit scoring using a fixed set of questions. The survey methodology and questionnaire align with previous similar research that analyses articles on credit scoring published in 1991–2015. We seek to compare our results with previous periods and highlight some of the recent best practices in the field that might be useful for future researchers. © 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). JEL classification: G320 Financing Policy; Financial Risk and Risk Management; Capital and Ownership Structure; Value of Firms; Goodwill; G210 Banks; Depository Institutions; Micro Finance Institutions; Mortgages; C440 Operations Research; Statistical Decision Theory; C650 Miscellaneous Mathematical Tools; C830 Survey Methods; Sampling Methods; C450 Neural Networks and Related Topics Keywords: Credit scoring; Survey; Statistics; Machine learning; Data mining; Performance assessment |
|  | @article{Zhang2023AnEC,  title={An ensemble credit scoring model based on logistic regression with heterogeneous balancing and weighting effects},  author={Runchi Zhang and Xue Liguo and Wang Qin},  journal={Expert Syst. Appl.},  year={2023},  volume={212},  pages={118732}, | An Ensemble Credit Scoring Model Based on Logistic Regression with Heterogeneous Balancing and Weighting Effects The logistic regression model is widely used in credit evaluation practice due to its strong interpretability of results, but its recognition performance for default samples which are minority in real-world imbalanced data sets need to be improved. This paper designs a novel ensemble model based on logistic regression as the logistic-BWE model. It first carries out data preprocessing, then applying sample balancing algorithm to generate several training sub data sets with different imbalance ratios and constructing sub models respectively, finally according to the performance of each sub model in the validation stage, the weight of different class of predicted results of each sub model is dynamically calculated. The empirical results indicate that compared with ten representative credit scoring models on six public data sets, the logistic-BWE model has the strongest ability to recognize default samples, and has the best generalization ability on most data sets while maintaining the interpretability of prediction results. Further tests demonstrate that the performance superiority of the logistic-BWE model is statistically significant, and it also has excellent robustness when it contains a sufficient number of sub models. Keywords：Logistic regression; Logistic-BWE model; Sample balancing algorithm; Ensemble credit scoring models; Dynamic weighting |
|  | @article{Yang2022BACSBA,  title={BACS: blockchain and AutoML-based technology for efficient credit scoring classification},  author={Fan Yang and Ya-nan Qiao and Yong Qi and Junge Bo and Xiao Wang},  journal={Annals of Operations Research},  year={2022},  pages={1 - 21}, | BACS: blockchain and AutoML-based technology for efficient credit scoring classification Credit evaluation is of high scientific significance and practical use, especially in today's plight of the world suffering from the COVID-19 epidemic. However, due to the difficulties inherent in credit scoring model building which involves a large number of data mining steps and requires a lot of time to process the data and build the model, efficient and accurate credit scoring methods are are urgently required. Aiming to solve this problem, we propose BACS, an blockchain and automated machine learning based classification model using credit dataset so that the credit modelling processes are performed in the pipeline in an automated manner to eventually obtain the classification results of credit scoring. BACS scheme consists of credit data storage to blockchain, feature extraction, feature selection, modelling algorithm and hyperparameter optimization, and model evaluation. Firstly, we propose a mechanism for credit data management and storage using blockchain to ensure that the entire credit scoring system is traceable and that the information of each scoring candidate is securely, efficiently and tamper-proofly stored on the blockchain nodes. Next, we design a pipeline using a random forest model to effectively integrate the key steps of credit data feature extraction, feature selection, credit model construction, and model evaluation. The experimental results demonstrate that our proposed automated machine learning-based credit scoring classification scheme BACS can assess the credit condition efficiently and accurately.  **Keywords:**Automated machine learning; Blockchain technology; Classification model; Credit crisis; Credit scoring; Hyperparameter optimisation. |
|  | @article{Wang2022DeepRL,  title={Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring},  author={Yadong Wang and Yanlin Jia and Yuhang Tian and Jin Xiao},  journal={Expert Syst. Appl.},  year={2022},  volume={200},  pages={117013}, | Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring Customer credit scoring is a dynamic interactive process. Simply designing the static reward function for deep reinforcement learning may be difficult to guide an agent to adapt to the change of the customer credit scoring environment. To solve this problem, we propose the deep Q-network with the confusion-matrix-based dynamic reward function (DQN-CMDRF) model. Especially, the new constructed dynamic reward function can adjust the reward dynamically according to the change of [confusion matrix](https://www.sciencedirect.com/topics/computer-science/confusion-matrix) after each deep Q-network model training, which can guide the agent to adapt to the change of environment quickly, so as to improve the customer credit scoring performance of the deep Q-network model. First, we formulate customer credit scoring as a finite [Markov decision process](https://www.sciencedirect.com/topics/computer-science/markov-decision-process). Second, to adjust the reward dynamically according to the customer credit scoring environment, the dynamic reward function is designed based on the [confusion matrix](https://www.sciencedirect.com/topics/computer-science/confusion-matrix). Finally, we introduce the confusion-matrix-based dynamic reward function into the deep Q-network model for customer credit scoring. To verify the effectiveness of the proposed model, we introduce four evaluation measures and make a series of experiments on the five customer credit scoring datasets. The experimental results show that the constructed dynamic reward function can more effectively improve customer credit scoring performance of the deep Q-network model, and the performance of the DQN-CMDRF model is significantly better than that of the other eight traditional classification models. More importantly, we find that the constructed dynamic reward function can accelerate the convergence speed and improve the stability of the deep Q-network model. |
|  | @article{Hayashi2022EmergingTI,  title={Emerging Trends in Deep Learning for Credit Scoring: A Review},  author={Yoichi Hayashi},  journal={Electronics},  year={2022}, | Emerging Trends in Deep Learning for Credit Scoring: A Review This systematic review aims to provide deep insights on emerging trends in, and the potential of, advanced deep learning techniques, such as machine learning algorithms being partially replaced by deep learning (DL) algorithms for credit scoring owing to the higher accuracy of the latter. This review also seeks to explain the reasons that deep belief networks (DBNs) can achieve higher accuracy than shallower networks, discusses the potential classification capabilities of DL-based classifiers, and bridges DL and explainable credit scoring. The theoretical characteristics of DBNs are also presented along with the reasons for their higher accuracy compared to that of shallower networks. Studies published between 2019 and 2022 were analysed to review and compare the most recent DL techniques that have been found to achieve higher accuracies than ensemble classifiers, their hybrids, rule extraction methods, and rule-based classifiers. The models reviewed in this study were evaluated and compared according to their accuracy and area under the receiver operating characteristic curve for the Australian, German (categorical), German (numerical), Japanese, and Taiwanese datasets, which are commonly used in the credit scoring community. This review paper also explains how tabular datasets are converted into images for the application of a two-dimensional convolutional neural network (CNN) and how “black box” models using local and global rule extraction and rule-based methods are applied in credit scoring. Finally, a new insight on the design of DL-based classifiers for credit scoring datasets is provided, along with a discussion on promising future research directions.  **Keywords:**  [**credit scoring**](https://www.mdpi.com/search?q=credit+scoring); [**credit risk**](https://www.mdpi.com/search?q=credit+risk); [**deep learning**](https://www.mdpi.com/search?q=deep+learning); [**convolutional neural networks**](https://www.mdpi.com/search?q=convolutional+neural+networks); [**tabular data**](https://www.mdpi.com/search?q=tabular+data); [**structured data**](https://www.mdpi.com/search?q=structured+data); [**deep belief networks**](https://www.mdpi.com/search?q=deep+belief+networks) |
|  | @article{Lenka2022EmpiricalAO,  title={Empirical Analysis of Ensemble Learning for Imbalanced Credit Scoring Datasets: A Systematic Review},  author={Sudhansu R. Lenka and Sukant Kishoro Bisoy and Rojalina Priyadarshini and Mangal Sain},  journal={Wireless Communications and Mobile Computing},  year={2022}, | Empirical Analysis of Ensemble Learning for Imbalanced Credit Scoring Datasets: A Systematic Review Credit scoring analysis has gained tremendous importance for researchers and the financial industries around the globe. It helps the financial industries to grant credits or loans to each deserving applicant with zero or minimal risks. However, developing an accurate and effective credit scoring model is a challenging task due to class imbalance and the presence of some irrelevant features. Recent researches show that ensemble learning has achieved supremacy in this field. In this paper, we performed an extensive comparative analysis of ensemble algorithms to bring further improvements in the algorithm oversampling, and feature selection (FS) techniques are implemented. The relevant features are identified by utilizing three FS techniques, such as information gain (IG), principal component analysis (PCA), and genetic algorithm (GA). Additionally, a comparative performance analysis is performed using 5 base and 14 ensemble models on three credit scoring datasets. The experimental results exhibit that the GA-based FS technique and CatBoost algorithm perform significantly better than other models in terms of five metrics, i.e., accuracy (ACC), area under the curve (AUC), F1-score, Brier score (BS), and Kolmogorov-Smirnov (KS). |
|  | @article{Hooman2016StatisticalAD,  title={Statistical and data mining methods in credit scoring},  author={Alireza Hooman and Govindan Marthandan and Wan Fauziah Wan Yusoff and Mohana Omid and Sasan Karamizadeh},  journal={The Journal of Developing Areas},  year={2016},  volume={50},  pages={371 - 381}, | [Statistical and data mining methods in credit scoring](https://www.semanticscholar.org/paper/Statistical-and-data-mining-methods-in-credit-Hooman-Marthandan/8c939302afce90f75a3068cc3f22ca7502eb0ed9)  The growing interest in the credit industry resulted in credit scoring being developed as an essential component, especially in the credit department of banks that deals with huge sums of credit data. When a bank or a credit corporation is assessing a credit application request, they will have to decide whether to approve or deny it. This necessitates the utilization of credit scoring. Although pioneers attempt to compensate for risks via interest rates, current investigations on financial conditions of different sections of society confirmed that interest could not replace risk assessment, which means that credit risk requires its own specialized assessment. With the assistance of sorting methods, credit scoring simplifies the decision-making process. It is almost impossible to analyze this large amount of data in the context of manpower and economy, although the data mining technique helps alleviate this complexity. Nowadays, there are a lot of data mining methodologies being utilized in the management of credit scoring. However, each method has its advantages and limitations, and there has not been a comprehensive approach in determining the most utilized data mining technique in the context of credit scoring. The major goal of this paper is to provide a complete literature survey on applied data mining methods, such as discriminant analysis, logistic regression, K-nearest neighbor, Bayesian classifier, decision tree, neural network, survival analysis, fuzzy rule-based system, support vector machine, and hybrid methods. These findings will assist researchers in realizing the most suitable approach in evaluating credit scores, pinpoint limitations, enhance them, and propose new approaches with improved capabilities. Finally, the limitations of the new approaches are discussed, and further suitable methods are recommended. |
|  | @article{Albareto2016DoesCS,  title={Does Credit Scoring Improve the Selection of Borrowers and Credit Quality?},  author={Giorgio Albareto and Roberto Felici and Enrico Sette},  journal={European Economics: Macroeconomics \& Monetary Economics eJournal},  year={2016}, | [Does Credit Scoring Improve the Selection of Borrowers and Credit Quality?](https://www.semanticscholar.org/paper/Does-Credit-Scoring-Improve-the-Selection-of-and-Albareto-Felici/af3aa3aa4d3ef689ca78f9bc8c0f39bcce4d92da)  This paper studies the effect of credit scoring by banks on bank lending to small businesses by addressing the following questions: does credit scoring increase or decrease the propensity of banks to grant credit? Does it improve the selection of borrowers? Does credit scoring improve or reduce the likelihood that a borrower defaults on its loan? We answer these questions using a unique dataset that collects data from both a targeted survey on credit scoring models and the Central Credit Register. We rely on instrumental variables to control for the potential endogeneity of credit scoring. We find that credit scoring does not change the propensity of banks to grant loans to the generality of borrowers but helps them select borrowers. We also find that credit scoring reduces the likelihood that a borrower defaults, in particular for smaller borrowers and for banks that declare to use credit scoring mainly as a tool to monitor borrowers. These results are homogeneous across bank characteristics such as size, capital, and profitability. Overall our results suggest that credit scoring has a positive effect on the selection of borrowers and on credit performance. |
|  | @inproceedings{Ha2016CreditSW,  title={Credit scoring with a feature selection approach based deep learning},  author={Van-Sang Ha and Ha-Nam Nguyen},  year={2016}, | [Credit scoring with a feature selection approach based deep learning](https://www.semanticscholar.org/paper/Credit-scoring-with-a-feature-selection-approach-Ha-Nguyen/50ddde187f34a555e9c565b7e046a354ac3c82af)  In financial risk, credit risk management is one of the most important issues in financial decision-making. Reliable credit scoring models are crucial for financial agencies to evaluate credit applications and have been widely studied in the field of machine learning and statistics. Deep learning is a powerful classification tool which is currently an active research area and successfully solves classification problems in many domains. Deep Learning provides training stability, generalization, and scalability with big data. Deep Learning is quickly becoming the algorithm of choice for the highest predictive accuracy. Feature selection is a process of selecting a subset of relevant features, which can decrease the dimensionality, reduce the running time, and improve the accuracy of classifiers. In this study, we constructed a credit scoring model based on deep learning and feature selection to evaluate the applicant’s credit score from the applicant’s input features. Two public datasets, Australia and German credit ones, have been used to test our method. The experimental results of the real world data showed that the proposed method results in a higher prediction rate than a baseline method for some certain datasets and also shows comparable and sometimes better performance than the feature selection methods widely used in credit scoring |
|  | @article{Chopra2018ApplicationOE,  title={Application of Ensemble Models in Credit Scoring Models},  author={Anjali Chopra and Priyanka Bhilare},  journal={Business Perspectives and Research},  year={2018},  volume={6},  pages={129 - 141}, | [Application of Ensemble Models in Credit Scoring Models](https://www.semanticscholar.org/paper/Application-of-Ensemble-Models-in-Credit-Scoring-Chopra-Bhilare/7ae71e83242b8ffbcb23413160e6eedfae94eec3)  Abstract Loan default is a serious problem in banking industries. Banking systems have strong processes in place for identification of customers with poor credit risk scores; however, most of the credit scoring models need to be constantly updated with newer variables and statistical techniques for improved accuracy. While totally eliminating default is almost impossible, loan risk teams, however, minimize the rate of default, thereby protecting banks from the adverse effects of loan default. Credit scoring models have used logistic regression and linear discriminant analysis for identification of potential defaulters. Newer and contemporary machine learning techniques have the ability to outperform classic old age techniques. This article aims to conduct empirical analysis on publically available bank loan dataset to study banking loan default using decision tree as the base learner and comparing it with ensemble tree learning techniques such as bagging, boosting, and random forests. The results of the empirical analysis suggest that the gradient boosting model outperforms the base decision tree learner, indicating that ensemble model works better than individual models. The study recommends that the risk team should adopt newer contemporary techniques to achieve better accuracy resulting in effective loan recovery strategies. |
|  | @article{Provenzano2020MachineLA,  title={Machine Learning approach for Credit Scoring},  author={Angela Rita Provenzano and Daniele Trifir{\`o} and Alessio Datteo and Lorenzo Giada and Nicola Jean and A. Riciputi and Giacomo Le Pera and Maurizio Spadaccino and Luca Massaron and Claudio Nordio},  journal={SSRN Electronic Journal},  year={2020}, | [Machine Learning approach for Credit Scoring](https://www.semanticscholar.org/paper/Machine-Learning-approach-for-Credit-Scoring-Provenzano-Trifir%C3%B2/1a9be760d058a890752423e6a122285cae629b9e)  In this work we build a stack of machine learning models aimed at composing a state-of-the-art credit rating and default prediction system, obtaining excellent out-of-sample performances. Our approach is an excursion through the most recent ML / AI concepts, starting from natural language processes (NLP) applied to economic sectors' (textual) descriptions using embedding and autoencoders (AE), going through the classification of defaultable firms on the base of a wide range of economic features using gradient boosting machines (GBM) and calibrating their probabilities paying due attention to the treatment of unbalanced samples. Finally we assign credit ratings through genetic algorithms (differential evolution, DE). Model interpretability is achieved by implementing recent techniques such as SHAP and LIME, which explain predictions locally in features' space. |
|  | @inproceedings{Kumar2020ReviewOM,  title={Review of Machine Learning models for Credit Scoring Analysis},  author={Madapuri Rudra Kumar and Vinit Kumar Gunjan},  year={2020}, | [Review of Machine Learning models for Credit Scoring Analysis](https://www.semanticscholar.org/paper/Review-of-Machine-Learning-models-for-Credit-Kumar-Gunjan/569766945340bb024fc818b24292dc045943d949)  Increase in computing power and the deeper usage of the robust computing systems in the financial system is propelling the business growth, improving the operational efficiency of the financial institutions, and increasing the effectiveness of the transaction processing solutions used by the organizations. Problem: Despite that the financial institutions are relying on the credit scoring patterns for analyzing the credit worthiness of the clients, still there are many factors that are imminent for improvement in the credit score evaluation patterns. There is need for improving the pattern to enhance the quality of analysis. Objective: Machine learning is offering immense potential in Fintech space and determining a personal credit score. Organizations by applying deep learning and machine learning techniques can tap individuals who are not being serviced by traditional financial institutions. Methodology: One of the major insights into the system is that the traditional models of banking intelligence solutions are predominantly the programmed models that can align with the information and banking systems that are used by the banks. But in the case of the machine-learning models that rely on algorithmic systems require more integral computation which is intrinsic. Hence, it can be advocated that the models usually need to have some decision lines wherein the dynamic calibration model must be streamlined. Such structure demands the dynamic calibration to have a decision tree system to empower with more integrated model changes. Results: The test analysis of the proposed machine learning model indicates effective and enhanced analysis process compared to the non-machine learning solutions. The model in terms of using various classifiers indicate potential ways in which the solution can be significant. Conclusion: If the systems can be developed to align with more pragmatic terms for analysis, it can help in improving the process conditions of customer profile analysis, wherein the process models have to be developed for comprehensive analysis and the ones that can make a sustainable solution for the credit system management. Originality: The proposed solution is effective and the one conceptualized to improve the credit scoring system patterns. If the model can be improved with more effective parameters and learning metrics, it can be sustainable outcome. Limitations: The model is tested in isolation and not in comparison to any of the existing credit scoring patterns. Only the inputs in terms of shortcomings from the existing models are taken in to account and accordingly the proposed solution is developed. |
|  | @article{Demajo2020ExplainableAF,  title={Explainable AI for Interpretable Credit Scoring},  author={Lara Marie Demajo and Vince Vella and Alexiei Dingli},  journal={ArXiv},  year={2020},  volume={abs/2012.03749}, | Explainable AI for Interpretable Credit Scoring  With the ever-growing achievements in Artificial Intelligence (AI) and the recent boosted enthusiasm in Financial Technology (FinTech), applications such as credit scoring have gained substantial academic interest. Credit scoring helps financial experts make better decisions regarding whether or not to accept a loan application, such that loans with a high probability of default are not accepted. Apart from the noisy and highly imbalanced data challenges faced by such credit scoring models, recent regulations such as the right to explanation' introduced by the General Data Protection Regulation (GDPR) and the Equal Credit Opportunity Act (ECOA) have added the need for model interpretability to ensure that algorithmic decisions are understandable and coherent. An interesting concept that has been recently introduced is eXplainable AI (XAI), which focuses on making black-box models more interpretable. In this work, we present a credit scoring model that is both accurate and interpretable. For classification, state-of-the-art performance on the Home Equity Line of Credit (HELOC) and Lending Club (LC) Datasets is achieved using the Extreme Gradient Boosting (XGBoost) model. The model is then further enhanced with a 360-degree explanation framework, which provides different explanations (i.e. global, local feature-based and local instance-based) that are required by different people in different situations. Evaluation through the use of functionallygrounded, application-grounded and human-grounded analysis show that the explanations provided are simple, consistent as well as satisfy the six predetermined hypotheses testing for correctness, effectiveness, easy understanding, detail sufficiency and trustworthiness. |
|  | @article{Laborda2021FeatureSI,  title={Feature Selection in a Credit Scoring Model},  author={Juan Laborda and Seyong Ryoo},  journal={Mathematics},  year={2021}, | [Feature Selection in a Credit Scoring Model](https://www.semanticscholar.org/paper/Feature-Selection-in-a-Credit-Scoring-Model-Laborda-Ryoo/5a673d120540c69807ae7355ed1b6c2c99b8f29e)  This paper proposes different classification algorithms—logistic regression, support vector machine, K-nearest neighbors, and random forest—in order to identify which candidates are likely to default for a credit scoring model. Three different feature selection methods are used in order to mitigate the overfitting in the curse of dimensionality of these classification algorithms: one filter method (Chi-squared test and correlation coefficients) and two wrapper methods (forward stepwise selection and backward stepwise selection). The performances of these three methods are discussed using two measures, the mean absolute error and the number of selected features. The methodology is applied for a valuable database of Taiwan. The results suggest that forward stepwise selection yields superior performance in each one of the classification algorithms used. The conclusions obtained are related to those in the literature, and their managerial implications are analyzed. |
|  | @article{Kozodoi2021FairnessIC,  title={Fairness in Credit Scoring: Assessment, Implementation and Profit Implications},  author={Nikita Kozodoi and Johannes Jacob and Stefan Lessmann},  journal={ArXiv},  year={2021},  volume={abs/2103.01907}, | [Fairness in Credit Scoring: Assessment, Implementation and Profit Implications](https://www.semanticscholar.org/paper/Fairness-in-Credit-Scoring%3A-Assessment%2C-and-Profit-Kozodoi-Jacob/da0bfa14e348c39d6ae104de6c576ec709f2e115)  The rise of algorithmic decision-making has spawned much research on fair machine learning (ML). Financial institutions use ML for building risk scorecards that support a range of credit-related decisions. Yet, the literature on fair ML in credit scoring is scarce. The paper makes three contributions. First, we revisit statistical fairness criteria and examine their adequacy for credit scoring. Second, we catalog algorithmic options for incorporating fairness goals in the ML model development pipeline. Last, we empirically compare different fairness processors in a profit-oriented credit scoring context using real-world data. The empirical results substantiate the evaluation of fairness measures, identify suitable options to implement fair credit scoring, and clarify the profit-fairness trade-off in lending decisions. We find that multiple fairness criteria can be approximately satisfied at once and recommend separation as a proper criterion for measuring the fairness of a scorecard. We also find fair in-processors to deliver a good balance between profit and fairness and show that algorithmic discrimination can be reduced to a reasonable level at a relatively low cost. The codes corresponding to the paper are available on GitHub1. |
|  | @article{Dastile2021MakingDL,  title={Making Deep Learning-Based Predictions for Credit Scoring Explainable},  author={Xolani Dastile and Turgay Çelik},  journal={IEEE Access},  year={2021},  volume={9},  pages={50426-50440}, | [Making Deep Learning-Based Predictions for Credit Scoring Explainable](https://www.semanticscholar.org/paper/Making-Deep-Learning-Based-Predictions-for-Credit-Dastile-%C3%87elik/3b3e8a3b16dec52a53a8e23df6947491cc564989)  Credit scoring has become an important risk management tool for money lending institutions. Over the years, statistical and classical machine learning models have been the most researched risk management tools in credit scoring literature, and recently the focus has turned to deep learning models. This transition is due to better performances that are shown by deep learning models in different domains. Despite deep learning models’ superior performances, there is still a need for explaining how these models make their predictions. The non-transparency nature of deep learning models has created a bottleneck for their use in credit scoring. Explanations of decisions are important for lending institutions since it is a requirement for automated decisions that are generated by non-transparent models to be explained. The other issue in using deep learning models, specifically 2D Convolutional Neural Networks (CNNs), in credit scoring is the need to have the data in image format. We propose an explainable deep learning model for credit scoring which can harness the performance benefits offered by deep learning and yet comply with the legislation requirements for the automated decision-making processes. The proposed method converts tabular datasets into images and thus allowing the application of 2D CNNs in credit scoring. Each pixel of the image corresponds to a feature bin of the tabular dataset. The predictions from the 2D CNNs were explained using state-of-the-art explanation methods. Furthermore, explanations were evaluated using a sanity check methodology and also performances of the explanation methods were compared quantitatively. The proposed explainable deep learning model outperforms the other credit scoring methods on publicly available credit scoring datasets. |
|  | @article{Wang2023CustomerCR,  title={Customer Credit Rating by Machine Learning},  author={Chengyijing Wang and Haining Jiang and Xiaoyan Jin and Ziyu Zhou},  journal={BCP Business \& Management},  year={2023}, | [Customer Credit Rating by Machine Learning](https://www.semanticscholar.org/paper/Customer-Credit-Rating-by-Machine-Learning-Wang-Jiang/28ed167caf95da7fc79fac6c684fe9486ca38be5)  Recently, people's consumption attitudes have also changed, being inclined to spend in advance. Banks and other financial institutions use credit rating models as a tool to evaluate the credit score of individuals, determine whether to grant the loan to the applicant. One of the biggest challenges for the banking industry in assessing the customers’ credit is that it is unlikely to provide a manual review to classify them because of the huge volume of data on applicants. Therefore, it is necessary to establish a suitable and effective credit rating model to help banks evaluate the quality of applicants. This paper focuses on the problems existing in the development of personal credit rating system and tries to find the best solution in the field of personal credit rating system. By selecting independent variables that are highly correlated with delinquency behavior, using different models for testing, and comparing the results of the models, this paper finally draws the conclusion that different algorithms combined by the group decision method can make better decisions. |
|  | @inproceedings{Amato2022CreditSP,  title={Credit Score Prediction Relying on Machine Learning},  author={Flora Amato and Antonino Ferraro and Antonio Galli and Francesco Moscato and Vincenzo Moscato and Giancarlo Sperl{\'i}},  booktitle={Sistemi Evoluti per Basi di Dati},  year={2022}, | [Credit Score Prediction Relying on Machine Learning](https://www.semanticscholar.org/paper/Credit-Score-Prediction-Relying-on-Machine-Learning-Amato-Ferraro/01bcb642be38e1217ae3a736473bc2f7c53b797d)  Financial institutions use a variety of methodologies to define their commercial and strategic policies, and a significant role is played by credit risk assessment. In recent years, different credit risk assessment services arose, providing Social Lending platforms to connect lenders and borrowers in a direct way without assisting of financial institutions. Despite the pros of these platforms in supporting fundraising process, there are different stems from multiple factors including lack of experience of lenders, missing or uncertain information about the borrower’s credit history. In order to handle these problems, credit risk assessments of financial transactions are usually modeled as a binary problem based on debt repayment, going to apply Machine Learning (ML) techniques. The paper represents an extended abstract of a recent work, where some of the authors performed a benchmarking among the most used credit risk assessment ML models in the field of predicting whether a loan will be repaid in a P2P platform. The experimental analysis is based on a real dataset of Social Lending (Lending Club), going to evaluate several evaluation metrics including AUC, sensitivity, specificity and explainability of the models |

1. GIỚI THIỆU

Chấm điểm tín dụng là quá trình đánh giá mức độ tín nhiệm của một cá nhân hoặc một tổ chức, các tổ chức tài chính phải sử dụng phương pháp chấm điểm tín dụng (1).Trong quá trình chấm điểm tín dụng, các yếu tố như lịch sử thanh toán, mức độ nợ nần, thu nhập và các thông tin tài chính khác được thu thập và phân tích để đánh giá khả năng thanh toán của cá nhân hoặc tổ chức, từ đó một số điểm tín dụng được gán cho khách hàng để công ty tài chính có thể dựa vào đó để đưa ra quyết định cung cấp tín dụng và điều chỉnh lãi suất. Trong thời đại ngày nay, chấm điểm tín dụng đã trở thành một yếu tố không thể thiếu trong ngành tài chính, đóng vai trò quan trọng trong việc đánh giá và quản lý rủi ro tín dụng- rủi ro tín dụng là rủi ro có tác động lớn nhất đối với bất kỳ ngân hàng và tổ chức tài chính nào (3). Phân tích chấm điểm tín dụng đã đạt được tầm quan trọng to lớn đối với các nhà nghiên cứu trong ngành tài chính, nó giúp cho tổ chức tài chính cấp tín dụng hoặc khoản vay cho những người nộp đơn xứng đáng với mức độ rủi ro tối thiểu hoặc bằng 0 (8). Sự quan tâm ngày càng tăng đối với ngành tín dụng dẫn đến việc chấm điểm tín dụng được phát triển như một phần thiết yếu (9). Trong một thị trường tài chính ngày càng phát triển nhanh chóng và phức tạp, các tổ chức tài chính ngày càng cần phải sử dụng các phương pháp phân tích tiên tiến để đánh giá rủi ro và đưa ra quyết định về việc cung cấp tín dụng (14). Sự phát triển của trí tuệ nhân tạo, học máy và phân tích dữ liệu đã mở ra cánh cửa cho việc áp dụng các mô hình dự đoán phức tạp hơn và hiệu quả hơn trong chấm điểm tín dụng. Việc chấm điểm tín dụng là cần thiết và rất quạ trọng. Mục tiêu của nghiên cứu này nhằm tìm ra phương pháp chấm điểm tín dụng để có thể giúp các tổ chức tài chính trong việc quyết định cấp tín dụng và quản lý rủi ro tín dụng.

Trong phần tiếp theo, chúng tôi giới thiệu một số nghiên cứu liên quan ở Phần II, Phương pháp đề xuất được đưa ra ở phần III. Phần IV là Thực nghiệm và kết quả, và Phần V là Kết luận.

1. CÁC NGHIÊN CỨU LIÊN QUAN

Cho đến nay nhiều nghiên cứu đã đề xuất các phương pháp chấm điểm tín dụng bằng nhiều mô hình khác nhau. Trong [1], nghiên cứu nói về một hệ thống chấm điểm tín dụng mới trong các tổ chức tài chính sử dụng công nghệ thông minh. Nghiên cứu cung cấp sự hiểu biết về quá trình ra quyết định đối với đặc điểm thông tin và chức năng của máy học (ML) trong các nhiệm vụ dự đoán tín dụng. Phân tích thực nghiệm cho thấy trong trường hợp tập dữ liệu của Đức, việc lựa chọn tính năng DT w GR và tối ưu hóa siêu tham số vượt trội hơn SVM NN với độ chính xác 99,78%. Đối với tập dữ liệu của Úc, SVM với lựa chọn tính năng G vượt trội hơn

Trong [2], bài nghiên cứu đề xuất một phương pháp mới kết hợp các ưu điểm của RF và SVM. RF dùng để trích xuất các đặc điểm từ tập dữ liệu và SVM được sử dụng vho phân loại bằng cách phân tích các thành phần và tác động của chúng đối với việc chấm điểm tín dụng. Nghiên cứu được xác thực dựa trên bộ dữ liệu rủi ro tín dụng của Đức, đạt độ chính xác cao tới 98,76%. Kết quả cho thấy rằng phương pháp đề xuất giúp đạt được kết quả tốt hơn cho phân tích điểm tín dụng.

Trong [4], nghiên cứu đề xuất mô hình chấm điểm tín dụng dựa trên hồi quy logistic với hiện ứng cân bằng và trọng số không đồng nhất. Kết quả thực nghiệm chỉ ra rằng so với các mô hình chấm điểm tín dụng đại diện trên sáu tập dữ liệu công cộng, mô hình logistic-B có khả năng nhận dạng mẫu mặc định mạnh nhất và có khả năng khái quát hóa b trên hầu hết các tập dữ liệu trong khi vẫn duy trì kết quả dự đoán diễn giải. Các thử nghiệm sâu hơn chứng minh rằng hiệu suất vượt trội của mô hình logistic-BWE có ý nghĩa thống kê và nó cũng có độ bền vượt trội khi chứa đủ số lượng mô hình phụ.

Trong [5], nghiên cứu này đề xuất BACS, một mô hình phân loại dựa trên Blockchain và máy học tự động sử dụng tập dữ liệu tín dụng để các quy trình lập mô hình tín dụng được thực hiện theo cách tự động để cuối cùng thu được kết quả phân loại của việc chấm điểm tín dụng. Các kết quả thử nghiệm chứng minh rằng sơ đồ phân loại tính điểm dựa trên máy học tự động được đề xuất BACS có thể đánh giá điều kiện tín dụng một cách hiệu quả và chính xác.